

## **MODELOS COMPETITIVOS PARA ANALISAR AS SÉRIES TEMPORAIS DA CONCENTRAÇÃO DO OXIGÉNIO DISSOLVIDO NO RIO VOUGA**

Magda Monteiro<sup>1,2</sup> e Marco Costa<sup>1,2</sup>

<sup>1</sup> ESTGA - Escola Superior de Tecnologia e Gestão de Águeda, Universidade de Aveiro

<sup>2</sup> CIDMA - Centro de Investigação e Desenvolvimento em Matemática e Aplicações, Universidade de Aveiro

### **RESUMO**

Neste estudo pretende-se avaliar o desempenho de modelos competitivos para descrever os valores mensais do oxigénio dissolvido em estações de monitorização da qualidade da água no rio Vouga. O confronto é realizado entre um modelo de regressão com erros correlacionados e dois modelos de espaço de estados.

**Palavras-chave:** Modelo de espaço de estados, filtro de Kalman, regressão linear, oxigénio dissolvido.

### **1. INTRODUÇÃO**

A avaliação da qualidade da água superficial é uma parte importante da monitorização ambiental, cuja avaliação pode prever a qualidade da água e evitar problemas de saúde pública de diversos tipos e níveis. Um papel importante na monitorização da qualidade da água superficial tem sido atribuído à variável concentração de oxigénio dissolvido (OD) uma vez que este indicador resulta do impacto de um conjunto de fatores ambientais como a temperatura da água, temperatura e pressão do ar, a morfologia do leito do rio, o estado de limpeza da água, as fontes de poluição das águas superficiais.

A análise química de OD mede a quantidade de oxigénio gasoso dissolvido numa solução aquosa. Para que a água seja considerada de boa qualidade é necessário que esta possua níveis adequados de OD, um elemento necessário para todas as formas de vida. Quando os níveis de OD na água descem abaixo de 5 mg/l, a vida aquática é colocada sob pressão. Níveis de oxigénio que permanecem abaixo de 1-2 mg/l, mesmo que por apenas algumas horas, podem resultar em grandes perdas de peixes ([5]). Assim, o estudo da evolução temporal da concentração do OD em diversos locais de um rio, bem como a predição de valores futuros são contributos importantes na monitorização da qualidade da água e na prevenção da poluição da mesma. Este trabalho apresenta um estudo comparativo entre vários modelos competitivos para descrever o comportamento mensal do OD na estação de monitorização de qualidade do Carvoeiro no rio Vouga. Uma descrição da bacia hidrográfica e caracterização das suas estações de monitorização pode ser encontrada no trabalho de [2].

O período usado na modelação compreende os meses de janeiro de 2002 a Maio de 2013 e os dados foram recolhidos do portal SNIRH ([6]). Estes modelos podem ainda ser avaliados quanto ao seu desempenho na predição de valores futuros do OD.

## 2. MODELOS EM ANÁLISE

Como a série da concentração de OD tem um comportamento periódico, os modelos que se apresentam são definidos tendo em conta essa característica da variável em estudo. O modelo base, que é comum aos modelos em estudo, é um modelo linear que tem em conta a variação sazonal na concentração de OD durante o ano, bem como a possibilidade da taxa de variação poder variar de acordo com o mês isto é, considera doze declives e doze ordenadas na origem, podendo ser escrito como

$$\begin{aligned} Y_t &= (\alpha_1 \cdot t + \beta_1)I_{t,1} + (\alpha_2 \cdot t + \beta_2)I_{t,2} + \dots + (\alpha_{12} \cdot t + \beta_{12})I_{t,12} + \xi_t \\ &= \sum_{s=1}^{12} (\alpha_s \cdot t + \beta_s) \cdot I_{t,s} + \xi_t, \end{aligned} \quad (1)$$

onde  $Y_t$  é a concentração do OD no mês  $t$ ,  $\alpha_s$  e  $\beta_s$ ,  $s = 1, 2, \dots, 12$ , são respetivamente o declive e a ordenada na origem associados ao mês  $t = s + 12k$ , para algum  $k = 0, 1, 2, \dots$ . A função indicatriz  $I_{t,s}$  é definida por  $I_{t,s} = 1$  if  $t = s + 12k$ , para algum  $k = 0, 1, 2, \dots$  e  $I_{t,s} = 0$  caso contrário, e  $\xi_t$  é um ruído branco ( $E(\xi_t) = 0$ ,  $var(\xi_t) = \sigma_\xi^2$  e  $E(\xi_t \xi_r) = 0$  for  $t \neq r$ ).

A estimação dos parâmetros do modelo, usando o método dos mínimos quadrados, revelou que, para o local em análise, os resíduos apresentavam uma correlação moderada. Adicionalmente, a análise dos resultados da estimação do modelo definido em (1), e após a correção dos erros padrão pelo facto dos erros serem correlacionados (ver [1]), apenas o declive associado ao mês de maio apresentou um valor estatisticamente significativo (para uma significância de 10%) pelo que o modelo base final possui 13 parâmetros,  $\alpha_5, \beta_1, \dots, \beta_{12}$ . Pretende-se então usar modelos que consigam captar melhor a correlação existente na série temporal do OD que têm na sua estrutura o modelo base de regressão com 13 parâmetros e a inclusão de outras componentes estocásticas que permitirão descrever melhor a série em estudo.

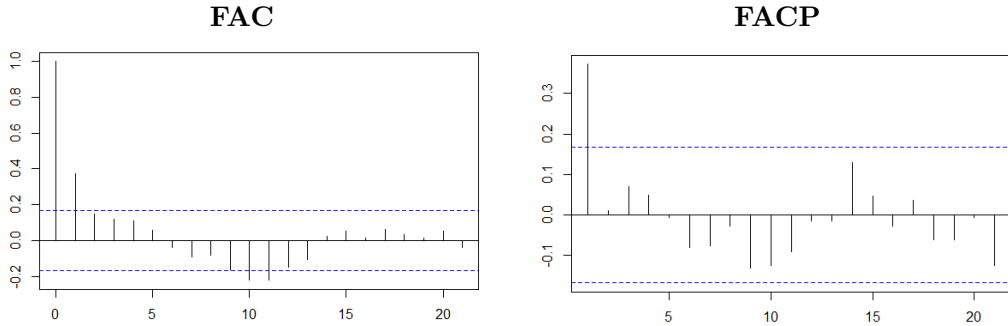


Figura 1: FAC e FACP empírica dos resíduos do modelo de regressão definido em (1).

### 2.1 Modelo de regressão linear com erros correlacionados (MI)

Tendo em conta que as funções empíricas de autocorrelação e de autocorrelação parcial que se apresentam na Figura 1, podemos definir o modelo linear com erros correlacionados, seguindo uma estrutura autorregressiva de ordem 1, descrito por

$$Y_t = \alpha_5 \cdot t \cdot I_{t,5} + \sum_{s=1}^{12} \beta_s \cdot I_{t,s} + \xi_t \quad (2)$$

$$\xi_t = \rho \xi_{t-1} + a_t, \quad (3)$$

com  $\{a_t\}$  um ruído branco cuja variância é  $\sigma_a^2$ .

A estimação dos parâmetros deste modelo pode ser efetuada por decomposição, estimando-se inicialmente os 13 parâmetros associados ao modelo base (1), pelo método dos mínimos quadrados, e posteriormente são estimados, pelo mesmo método, os parâmetros da equação (3), usando os resíduos obtidos na primeira etapa de estimação.

## 2.2 Modelo em espaço de estados aditivo (MII)

Atendendo a que a variabilidade associada ao processo autorregressivo no modelo MI poderá não ser suficiente para explicar toda a variabilidade da série de OD, considera-se o modelo mais geral, que contém duas fontes de variabilidade. O modelo de espaço de estados é definido por

$$Y_t = \alpha_5 \cdot t \cdot I_{t,5} + \sum_{s=1}^{12} \beta_s \cdot I_{t,s} + \xi_t + \epsilon_t \quad (4)$$

$$\xi_t = \rho \xi_{t-1} + a_t. \quad (5)$$

As equações (4)–(5) são respetivamente as equações de observação e de estado. Na equação de observação, os valores observados dependem de uma componente conhecida, de um fator estocástico aditivo  $\{\xi_t\}$  e uma componente de erro aleatória,  $\{\epsilon_t\}$ , designada por processo dos erros de observação que são descritos por um ruído branco com variância  $\sigma_\epsilon^2$ . A equação do estado descreve o comportamento da componente estocástica,  $\{\xi_t\}$ , que se assume ser um processo autorregressivo de ordem 1 estacionário de média zero e parâmetro autorregressivo  $|\rho| < 1$ ; o erro do estado,  $\{a_t\}$  é não correlacionado com o processo dos erros de observação ( $E(a_t \epsilon_r) = 0, \forall t, r$ ) e também descrito por um ruído branco. Adicionalmente os processos de ambos os erros são considerados ter distribuição normal. De notar que o modelo MI é um caso particular deste modelo quando se admite que a variância associada ao erro de observação é nula.

## 2.3 Modelo de calibração (MIII)

O terceiro modelo possui também duas componentes principais numa estrutura multiplicativa. A primeira é a componente base de regressão e a segunda componente estocástica que mensalmente irá calibrar a primeira componente. O modelo acima descrito é um modelo em espaço de estados, que pode ser definido por

$$Y_t = X_t \cdot \left( \alpha_5 \cdot t \cdot I_{t,5} + \sum_{s=1}^{12} \beta_s \cdot I_{t,s} \right) + e_t \quad (6)$$

$$X_t = \mu + \phi(X_{t-1} - \mu) + \varepsilon_t. \quad (7)$$

Este modelo difere do anterior por ter uma estrutura multiplicativa, em que processo do estado  $\{X_t\}$  tem uma estrutura autorregressiva de ordem 1 de média não nula  $\mu$ . Também neste modelo os processos dos erros de observação e de estado são ruídos brancos e não correlacionados entre si. A distribuição associada aos erros é assumida ser a distribuição Gaussiana. Nos modelos MII e MIII, devido à sua estrutura, a estimação dos parâmetros destes modelos podem ser realizada em duas etapas. Na primeira etapa, que é comum aos três modelos, são estimados os parâmetros do modelo de regressão base, que, numa fase posterior, são considerados como conhecidos e serão usados para estimar os parâmetros associados ao estado. Nesta segunda etapa, como o estado é não observável e os modelos são lineares em que se assume a normalidade dos erros de observação e de estado, a estimação dos parâmetros é realizada através do método de máxima verosimilhança em que a predição a um passo dos valores observados de  $Y$  (ou de uma série obtida a partir desta) é obtida pelo algoritmo do filtro de Kalman ([4],[3]). Para a obtenção das estimativas dos parâmetros de máxima verosimilhança é necessário a utilização de procedimentos numéricos disponíveis em diversos *softwares*.

### 3. ALGUNS RESULTADOS

A Tabela 1 resume os resultados da modelação do OD na estação do Carvoeiro, rio Vouga no período em análise. Os erros quadráticos médios dos três modelos estão próximos, assim como os coeficientes de determinação. Os modelos MI e MII têm um desempenho ligeiramente superior ao modelo MIII. De salientar que a estimativa da variância do erro de observação do modelo MII é estatisticamente significativa, diferenciando assim os dois modelos MI e MII. No que concerne à análise dos resíduos dos modelos em estudo, estes não apresentam correlação significativa e os histogramas não estão visualmente longe da normalidade. No entanto apenas no modelo MII o teste de Kolmogorov-Smirnov não rejeita a normalidade dos resíduos que é um pressuposto comum a todos os modelos.

Tabela 1: Resultados da estimação dos parâmetros dos modelos

Modelo Base											
par.	est.	err. pad.*									
$\alpha_5$	-0.010	0.005									
$\beta_1$	10.209	0.25									
$\beta_2$	9.983	0.24									
$\beta_3$	9.546	0.24									
$\beta_4$	8.975	0.44									
$\beta_5$	9.262	0.24									
$\beta_6$	7.771	0.25									
$\beta_7$	8.164	0.25									
$\beta_8$	8.844	0.25									
$\beta_9$	7.941	0.25									
$\beta_{10}$	7.844	0.25									
$\beta_{11}$	8.968	0.25									
$\beta_{12}$	9.946	0.25									
$R^2$ 0.561			$R^2$ 0.598			$R^2$ 0.602			$R^2$ 0.595		
E.Q.M. 0.579			E.Q.M. 0.528			E.Q.M. 0.528			E.Q.M. 0.534		

### AGRADECIMENTOS

Os autores foram parcialmente financiados por fundos portugueses através do CIDMA e da FCT-Fundação para a Ciência e a Tecnologia, dentro do projeto UID/MAT/04106/2013.

### Referências

- [1] Alpuim, T. e El-Shaarawi, A. (2008) On the efficiency of regression analysis with AR(p) errors. *Journal of Applied Statistics*, 35:7, 717–737.
- [2] Costa, M; Monteiro, M. (2016). Discrimination of water quality monitoring sites in River Vouga using a mixed-effect state space model. *Stochastic Environmental Research and Risk Assessment* 30, 2: 607–619.
- [3] Costa, M, Monteiro, M. (2016). Bias-correction of Kalman filter estimators associated to a linear state space model with estimated parameters. *Journal of Statistical Planning and Inference*, 176: 22 - 32.
- [4] Harvey AC. (2006). *Forecasting structural time series models and the Kalman filter*. Cambridge, Cambridge University Press.
- [5] Shifflett DS. Water and Sustainability. <http://www.unc.edu/~shashi/TablePages/dissolvedoxygen.html> (acedido a 16 de julho de 2014).
- [6] Portal da Água. Instituto da Água. I.P. (INAG). <http://portaldaagua.inag.pt> (acedido a 20 Março de 2018).